



Recommender project for opening a book café business in Bogotá, Colombia

July, 2021



Objective

- Determine the best location to open a new business in Bogota, Colombia
- Capture data of Bogota's Localities, Neighborhoods and its venues
- Capture data of covid cases in Bogota so as to take this information into account for the final location decision
- Run a clusterization process to determine which neighborhoods are most likely to have the sort of business we are focusing on in the vicinity

Possible Stakeholders

- People interested in opening a business now that some covid restrictions are being lifted and the economy is being reestablished
- People interested in opening a business in the book café category



Data Acquisition

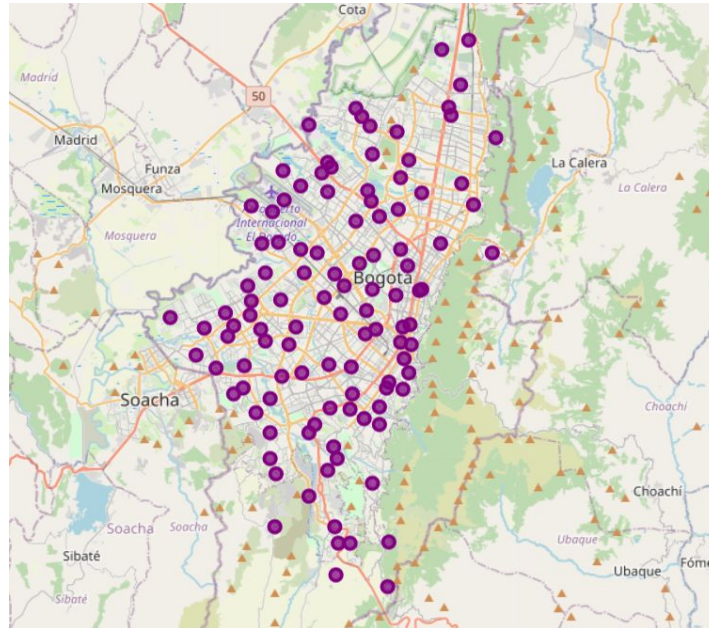
- Bogota's Localities and Neighborhoods. Taken from [https://es.wikipedia.org/wiki/Unidades de Planeamiento Zonal](https://es.wikipedia.org/wiki/Unidades_de_Planeamiento_Zonal)
- Bogota's Venues and HotSpots. Captured by using the FourSquare API
- Bogota's Localities and Neighborhood coordinates. Captured by using the Geopy Python Libraries
- Bogota's Covid Cases 2020/2021. Taken from <https://datosabiertos.bogota.gov.co/dataset/numero-de-casos-confirmados-por-el-laboratorio-de-covid-19-bogota-d-c/resource/b64ba3c4-9e41-41b8-b3fd-2da21d627558>



Data Cleaning

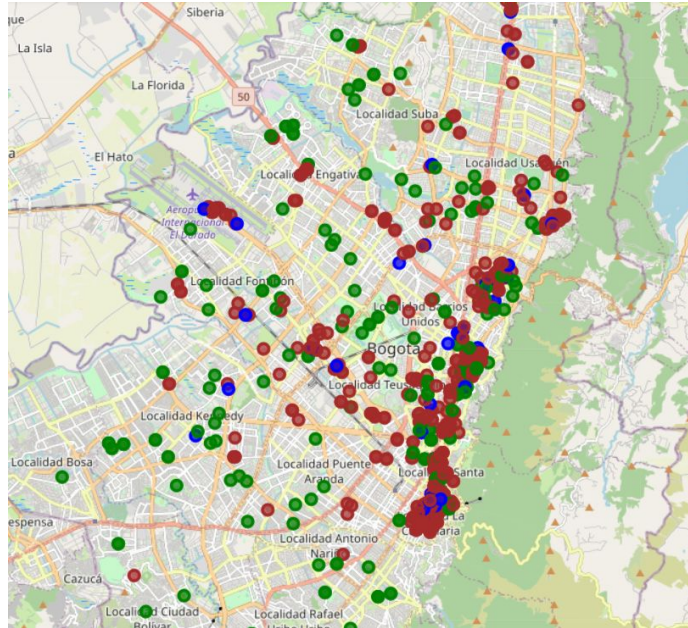
- Change the names of neighborhoods and localities so that these can match between dataframes
- Change type of variables to their correct format (e.g Covid Diagnosis Date from object to date)
- Drop records which are not relevant to the project (e.g Covid cases outside of Bogota)
- Drop columns and feature which are not relevant to the project

Data Exploration - Bogota's Neighborhoods Location

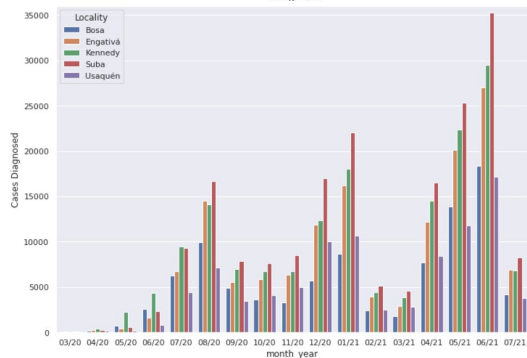
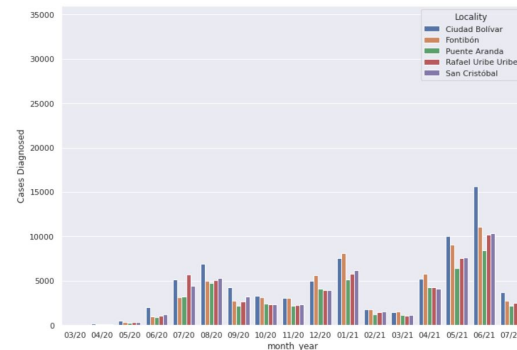
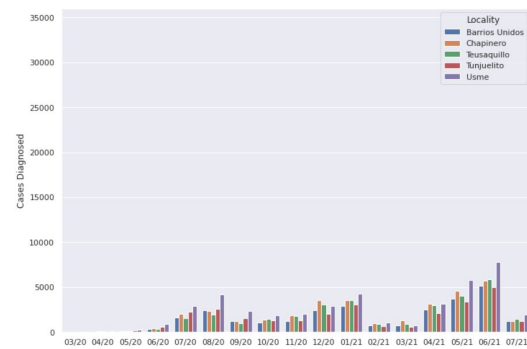
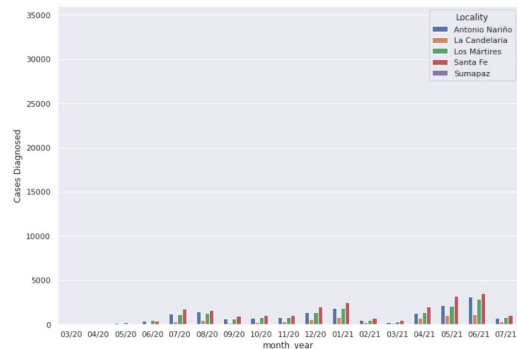


Data Exploration - Venues

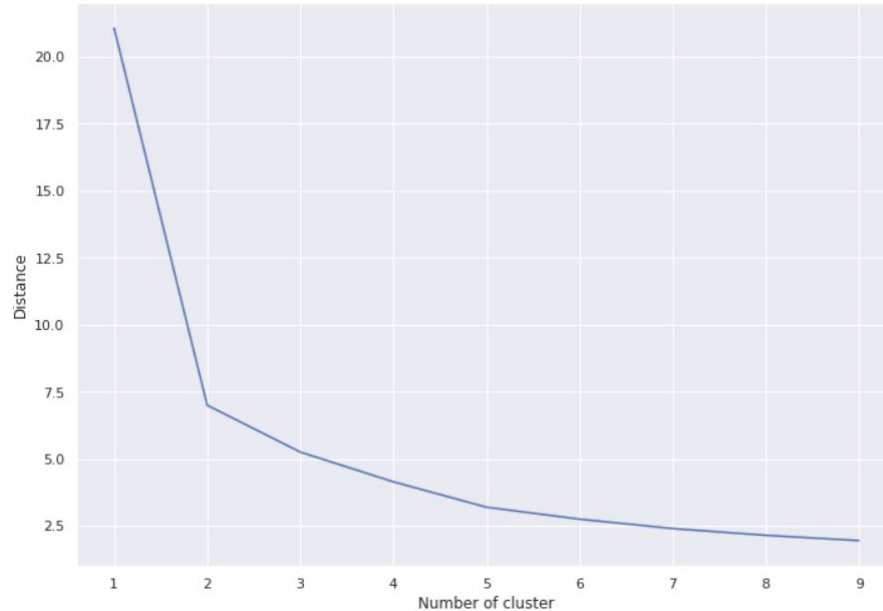
Most target businesses seem to be located at the eastern side of the city



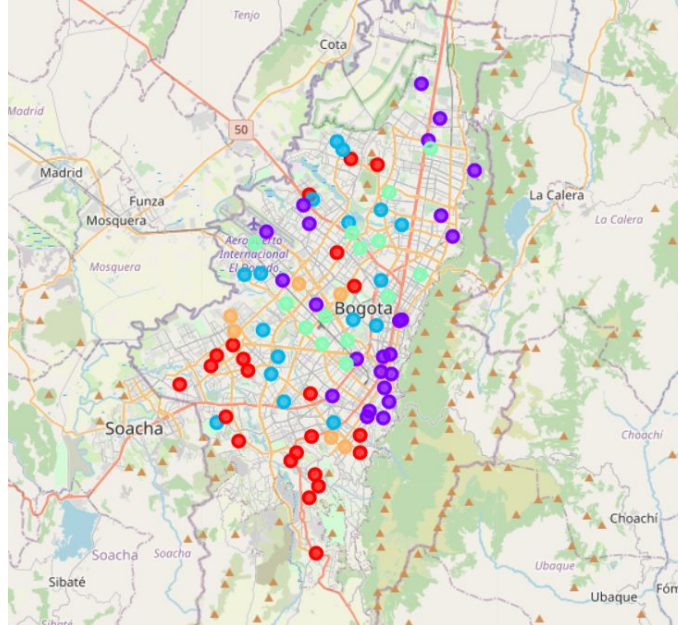
Data Exploration - Covid Cases



K-Means Clustering - Elbow method for # of clusters

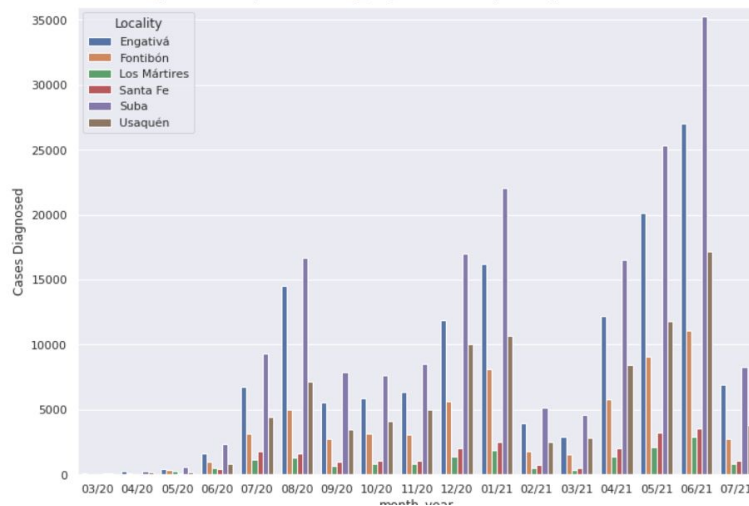


K-Means Clustering - Map with Neighborhood Clusters



Recommendation - Neighborhoods and Covid Cases

Neighborhood	Coffee Shops	Book Shops	Parks	Locality
San Cristóbal Norte	1.0	NaN	NaN	usaquén
La Academia	1.0	NaN	NaN	suba
Las Cruces	2.0	NaN	1.0	santa fe
La Uribe	3.0	1.0	NaN	usaquén
Capellanía	3.0	1.0	1.0	fontibón
Santa Isabel	3.0	NaN	1.0	los mártires
Bavaria	6.0	NaN	NaN	kennedy
Garcés Navas	6.0	NaN	2.0	engativá
Verbenal	7.0	1.0	1.0	usaquén
Country Club	7.0	1.0	2.0	usaquén
Ciudad Salitre Occidente	7.0	1.0	2.0	fontibón





Conclusions

- K-means Clustering is a good algorithm for this sort of project and for the data that we had available
- The Foursquare API is useful to get venues from different cities but it is possible not all venues are available through the API calls
- The project has room for new versions were there could be more automation of certain functions and to allow the client to choose from a wider range of businesses categories